ELSEVIER
Available online at www.sciencedirect.com
ScienceDirect
Procedia Engineering 7 (2010) 356–359
Procedia Engineering
www.elsevier.com/locate/procedia

2010 Symposium on Security Detection and Information Processing

An Improved Attribute Reduction Algorithm Based on Importance of Attribute Value

Ju Li [a*], Xiaowen Fan [b], Xing Wang [c]

[a] *School of Computer Science and Engineering, Chang Shu Institute of Technology, Changshu, Jiangsu, 215500, China*
[b] *School of Computer Science and Engineering, Chang Shu Institute of Technology, Changshu, Jiangsu, 215500, China*
[c] *School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, 210044, China*

Abstract

The theory of rough set is a new mathematical tool to deal with the uncertain problems, and the attribute reduction is one of key problems in the theory. To obtain a better method of attributes reduction in decision-making system, a concept of restrictive positive region was proposed. The method using positive region and restrictive positive region which have been obtained, shrinks the scope of data processing, so that reduces the time demand. By an instance, the paper presents the application of this method, and confirms that the computation is reduced and the result could be simpler, using this algorithm compared with the traditional algorithm. Thus, it is proven that this method is a fast and efficient algorithm of attribute reduction.

Keywords: rough set; attribute reduction; restrictive positive region; heuristic algorithm

1. Introduction

The main idea of rough set theory [1] [2] is that, uses the known-knowledge from the knowledge base to describe approximately the knowledge which is inaccurate or uncertain, under the premise of maintaining the classification of information systems without changing, and export the problems of decision-making or classification rules by adding knowledge and reduction. The key problem of this theory is the attribute reduction [3] [4] and the rule reduction [5]. An algorithm of heuristic rule reduction by a relatively positive region as additional information was presented by Zhu Hong [6]. The method of property value reduction, by considering its relative positive region whether will be changed when removing the value of each attribute in turn. Since the algorithm needs to be considered for each attribute value, it is more complex. The algorithm of Core searching proposed by KarnoBozi [7] is simple, fast and has better effect for attribute reduction. But the Core Searching algorithm needs to search all the remaining items in the discernibility matrix, and will continue until the matrix is empty when search for a candidate attribute reduction. Therefore, the algorithm will not only increase some additional number of searching count, but also raise a large amount of calculation, and the resulting reduction set will also have property redundancy items. This paper proposed

[*] Corresponding author. Tel.:+86-0512-52166136.
E-mail address: liju284532@163.com.

1877-7058 © 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2010.11.057

an improved attribute reduction algorithm, by research on the algorithm about the importance of various attributes. The method using positive region and restrictive positive region which have been obtained, shrinks the scope of data processing, so that reduces the time demand. And the experimental results confirmed the feasibility of the algorithm.

2. The basic concept of rough set

Definition 1: For a given decision-making system $S = (U, C \cup D, V, f)$, reduction of condition attribute set C is a non-empty subset of C-P. It meets：

① $\forall a \in P$, cannot be omitted by D

② $POS_p(D) = POS_c(D)$

Claimed: P is a reduction of C. The set of all reduction of C denoted RED(C).

By the reduction of the definition, every decision-making system reduction may have several, but reduction is equivalent, that is say they have the same classification ability. The reduction of nuclear is the most important attribute set, which includes all of the reduction. Definition 2：For a given decision-making system $S = (U, C \text{ Y } D, V, f)$, C is condition attribute set, D is decision-making attribute set, $E \subseteq C$，$c \in E$，When removed c from the property E, property c defined on the importance of D in E：

$$SIG_E^D(c) = \left| POS_E(D) - POS_{E-\{c\}}(D) \right|$$

When adding attribute c to the E, the attribute c in the expansion of E defined on the importance of D：

$$SIG^D_{E\cup\{c\}}(c) = \left| POS_{E\cup\{c\}}(D) - POS_E(D) \right|$$

The definition says the importance of an attribute relative to the decision attribute can be used to remove the property from the property after the focus is formed to determine the domain size changes, When SIGDG(c)＞0 explain c cannot reduction, When SIGDG(c)=0, explain attribute c is redundant, the greater SIGDG(c) is Note after removing c is the domain size of the changes caused by the greater, That is, the correct classification of the decision attribute can be divided into equivalence classes of elements in the less, c on the more important decision attribute.

Definition 3: In the knowledge representation system $S = (U, C \cup D, V, f)$中, $U_1 \subseteq U$, U_1≡Empty set, the C domain of D is limited to U_1 $SIG_C^{u_1}(D) = \underset{X \in U_1 / D}{\text{Y}} C\_(X)$, Referred to as the restrictions are on the U_1 domain.

Algorithm basic steps are as follows:

Step 1: Calculate C positive region of the decision attribute D, $POS_c(D)$; make CORED (C)= empty set, Calculated each attribute $c \in C$ In C, the importance of the D,SIGDC(C), If SIGDC(C)≠0, then CORED(C)= CORED(C)∪{c}, Finally we get CORED(C) was C relative to D for the relative nuclear. make A=CORED(C)

Step 2：If $POS_A(D)$=POSC(D), terminate computation. CORED(C) is C relative to D, a relative reduction; otherwise, to step three.

Step 3：make I=1, U_1=U-POSA(D).

Step4、From I = 1 to begin a subset of condition attributes C—A repeat：

(1) For each attribute $c \in C - A$, Calculate the positive region on UI after properties c added A , $POS^{U_I}{}_{A\text{Y}\{c\}}(D)$ $\left| POS^{U_I}{}_{A\text{Y}\{c\}}(D) \right|$ is 0, Then put out reduction;

(2) Find $\left| POS^{U_I}{}_{A\text{Y}\{c\}}(D) \right|$ the set of the largest property c-T;

(3) Calculate for the branch each attribute q in T, $A = A \text{ Y } \{q\}$;

(4) I=I+1, calculate $POS_A(D)$ and $U_I = U_{I-1} - POS^{U_{I-1}}{}_A(D)$;

IF($POS_A(D) = POS_C(D)$) { put out C relative to D, a reduction, and $T = T - \{q\}$; if T is not empty, turn(3); Otherwise, turn step 5 }

ELSE turn (1)

Step 5：Finally get all the relative reduction.

3. Algorithm analyses

This paper analyses the superiority of the improved algorithm by an example. See the following example：

	Condition attribute			Decision attribute
On the Field	a	b	c	d
1	1	1	3	2
2	1	1	1	1
3	1	1	0	1
4	2	2	1	2
5	2	2	1	1

Step 1：U/C={{1}, {2}, {3}, {4, 5} }, U/D={{2, 3, 5}, {1, 4} }, $POS_{(C)}$ (D) ={{1}, {2}, {3} }; U/(C-{a})={{1}, {2}, {3}, {4, 5}}, U/(C-{b}) ={{1}, {2}, {3}, {4, 5}}, U/(C-{c}) ={{1, 2, 3}, {4, 5} }, U/D={{1, 4 }, {2, 3, 5}}, POS(C-{a}) (D) = POS(C-{b}) (D) = POS(C) (D) ={{1}, {2}, {3} } , $POS\{c\}(D) = \phi$. Because,

$SIG_C^D(c) = \left|POS_C(D) - POS_{C-\{c\}}(D)\right|$, so,

$SIG_C^D(a) = SIG_C^D(b) = 0, SIG_C^D(c) \neq 0$, so,

$A = CORE_D(C) = \{c\}$;

Step 2：U/A=U/{c} = {{2, 4, 5}, {1}, {3}}, POS_A (D) = {{1}, {3}} $\neq$ $POS_{(C)}$ (D), so, A is not the smallest reduction;

Step 3：make I=1, U_1 =U- POS_A(D) ={{2}, {4}, {5}};

Step 4：C-A= {a, b},

(1) $U1 /D = \{\{2, 5\}, \{4\}\}, U_1 /A \cup \{a\} = \{\{1\}, \{2\}, \{3\}, \{4, 5\}\}$,

$U_1 /A \cup \{b\} = \{\{1\}, \{2\}, \{3\}, \{4, 5\}\}$, $POS^{U_1}{}_{A\cup\{a\}}(D) = POS^{U_1}{}_{A\cup\{b\}}(D) = \{\{1\}, \{2\}, \{3\}\}$, so the degree of importance of {a}, {b} is the same, we take T={a, b}.

(2) First we consider a, $A = A \cup \{a\} = \{a, c\}$, $POS^{U_1}{}_A(D) = \{\{2\}, \{4, 5\}\}$, $U/A = U/(C - b)$,

$POS_A(D) = POS_{C-b}(D) = POS(C)(D)$, so, RED_1={a, c} is one of reduction.

(3) Second, consider the b, $A = A \cup \{b\} = \{b, c\}$, $POS^{U_1}{}_A(D) = \{\{2\}, \{4, 5\}\}$, U/A=U/(C-a),

$POS_A(D) = POS_C - a(D) = POS(C)(D)$, so, RED_2={b, c} is also one of the reduction.

Through the above analysis shows that, the new algorithm to search optimal or sub optimal reduction is superior to the traditional algorithms.

4. Conclusions

The paper took into account the problem of positive domain and boundary domain, and present a measurement method based on attribute importance, and constructed “$POS^{U_1}{}_{A\cup\{c\}}(D)$”. The experimental results show that the algorithm is better than only the traditional algorithm only using positive region as Heuristic information. In order to find the minimum conditions for attribute reduction algorithm, the new method reduces the search space and computation time.

References

[1] Pawlak Z. Rough sets. *Int. J.comput.Inf.Sci.* 1982; **11**:341-356.
[2] Pawlak Z. Rough classification. *Int.J. Man Mach. Stud.* 1984; **20**: 469-483.

[3] Chang Liyun. Rough set attribute reduction and algorithm and its application. Chongqing University of Posts and Telecommunications Master thesis, 1999.
[4]Zhi Tianyun, Miao Duoqian. Binary discernible matrix of transformation and efficient attribute reduction algorithm for. *Computer Science* 2002; **2**: 140-142.
[5] Hao Lina, Xu Xinhe. Rules correction method based on rough set theory. *Control and Decision* 2002; **5**: 614-616.
[6] Zhu Hong. an algorithm based on property and property value reduction of Rough Set [J]. 2002; **24 (3)**: 36239.
[7] Karno Bozi. Core searching on rough sets[C]. Proceedings of the 23rd Int Conf Information Technology Interface. Pula, Croatia: IEEE Press, 2001; 217-222.